ASTROCHEMICAL FORECASTING WITH MACHINE LEARNING

KELVIN LEE, *Department of Chemistry, MIT, Cambridge, MA, USA*; ANDREW M BURKHARDT, *Smithsonian Astrophysical Observatory, Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA*; MICHAEL C McCARTHY, *Atomic and Molecular Physics, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA*; BRETT A. McGUIRE, *Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*.

Since the first molecules were detected in space, we have now reached a point where chemical and physical complexity in the interstellar medium reaches the boundaries of what human expertise and intuition alone can achieve. With every new molecule we discover, the question "What comes next?" grows more and more difficult to answer as more possibilities emerge. Conventionally, we turn to chemical models for guidance; this may be complicated when considering complex, non-LTE processes such as shocks, radiation, and grain-surface chemistry. Moreover, expansion of chemical networks typically requires hand-picked reactions and species, requiring an exhaustive knowledge of chemical and astrophysical literature, and can impose human bias on which reactions and molecules are important. As a complimentary approach to conventional chemical models, we have developed an unsupervised machine learning pipeline for predicting molecular abundances in a non-parametric fashion. Leveraging tools originally developed in high throughput drug discovery and data science, our pipeline captures and uses millions of molecules from various databases to create chemically descriptive vector representations for quantitative comparison. These representations are subsequently used to predict molecular properties in a given environment; as a proof-of-concept, we use the well-characterized chemical inventory of TMC-1, including the latest discoveries from the GOTHAM collaboration. We show that the model can be successfully conditioned on an inventory, able to reproduce column densities of unseen molecules to within an order of magnitude without any tuning parameters. Simultaneously, we are able to use the model to predict column densities of hundreds of thousands of molecules not yet detected in space, as a way to guide efforts, as well as provide a robust statistical baseline for expected abundances.